DOCUMENT RESUME

ED 083 285 TM 003 248

AUTHOR Messick, Samuel
TITLE The Context of Assessment and the Assessment of Context.
NOTE 7p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Early Childhood Education; Evaluation Techniques; *Interaction Process Analysis; Participant Characteristics; *Program Evaluation; *Student Testing; Technical Reports
IDENTIFIERS *CIRCUS Assessment Measures

ABSTRACT

To measure any element or characteristics of an early childhood education system, the general context of interdependencies must be assessed in order to take into account possible interactions of the characteristics measured with characteristics of the student, teacher, situation, and background. A comprehensive program of individual assessment should include provision for gauging three major aspects of context: (1) inferences about personal characteristics, particularly about competencies, should be relative to the context of environment, education experiences, and programs to which the child has been exposed; (2) inferences about a particular characteristic or competency should be relative to the context of his general personality and intellectual makeup, or at least to the salient features of that makeup; and (3) inferences about measured characteristics should be relative to the context of the measurement process per se. Strategies for the assessment of these aspects of context, particularly as exemplified in the ETS CIRCUS approach to comprehensive assessment, are considered. For a comprehensive program of measurement to deal meaningfully with the assessment of context, it must include provision for multivariate analysis and for the display, reporting, and interpretation of interactive and moderated relationships. (DB)

FILMED FROM BEST AVAILABLE COPY

ED 083285

TM 003 248

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Context of Assessment and the Assessment of Context

Samuel Messick
Educational Testing Service

Early childhood education is an extremely complicated system--it involves, at the very least, a set of complex, multifaceted organisms changing over time in interaction with diverse environmental influences. Furthermore, this system is composed of differentiated but overlapping subsystems that embrace the child, family, community, and various peer groups as well as the school, teachers, and programs. Since the concept of system implies a functioning whole whose various elements and subsystems are interdependent, it follows that the operation of one part of the system may interact with and produce unanticipated consequences in other parts of the system.

In attempting to measure any element or characteristic of such a system, it is necessary to assess the general context of interdependencies in order to take into account possible interactions of the characteristics measured with student, teacher, situation, and background characteristics. Otherwise we are at a loss to know how to generalize the measure and its meaning (or to limit its generalization) across student groups and across situations.

This relativity of inferences about measured characteristics to context has three major aspects: First, inferences about personal characteristics, particularly about competencies, should be relative to the context of environment, educational experiences, and programs to which the child has been exposed. When inferences about competency are drawn from test performance, it should make a difference whether or not the child has had an opportunity to

learn the skills required by the task or whether the child (or his teachers or parents or peers) thought those skills were important or relevant. Second, inferences about a particular characteristic or competency of a child should be relative to the context of his general personality and intellectual makeup, or at least to the salient features of that makeup. The child himself is a very complicated system of interdependencies, and one must anticipate that certain of his traits and characteristics will influence or interfere with the assessment of other traits and characteristics. Third, inferences about measured characteristics should be relative to the context of the measurement process per se--not just by taking into account critical objective features, such as whether the task was timed or untimed, but by tempering interpretations of test responses in light of the child's general style of reaction to the task, the tester and the testing situation.

A comprehensive program of individual assessment should include provision for gauging, even if only in rudimentary fashion, these three major aspects of context, for if we are sensitive to the issues, even relatively primitive indicators of contextual interactions can have a profound influence on interpretative practice. They can provide warning signals, for example, that certain generalizations may be unwarranted, that alternative hypotheses should be seriously entertained, or that additional measurement should be undertaken to clarify ambiguities.

Let us consider some strategies for the assessment of these major aspects of context, particularly as exemplified in the ETS CIRCUS approach to comprehensive assessment.

I. Environmental and program context is perhaps ideally assessed through direct observation using multiple independent observers, but it may also be conveniently and much less expensively assayed using indigenous, though biased, observers by means of a teacher questionnaire. Since teachers are prime agents in the educational context afforded the child, their biases are important to document in their own right, and a teacher questionnaire offers a ready means not only for eliciting teachers' descriptions of class and program characteristics, but also for appraising attitudes and viewpoints that might influence both their judgment and their teaching behavior.

Through this questionnaire mode, then, teachers are asked to describe the background of each child in their class in terms of age, sex, ethnic group membership, family occupational status, and previous educational experience; to describe the structure and setting of the classroom, the materials and facilities available along with the extent of their utilization, and the relative amounts of a variety of classroom activities; and to characterize briefly the school or center of which the class is a part. In addition, the teachers are asked several questions about previous experience and education, job attitudes and preferences, educational viewpoints, and predilections for various educational techniques and objectives.

This direct questioning of teachers about their programs and preferences may draw their attention to gaps in desirable facilities and activities or to an underemphasis upon valuable techniques and objectives, and these imbalances may come to be redressed in subsequent practice. This may be all to the good educationally, but we should be sensitive to the possibility that such a reactive approach to the assessment of context may be obtrusive and

hence may change or distort the very context it is meant to assess. From a research standpoint, this is an interesting but possibly minor caveat. It points to one out of many possible sources of reliable change in context and, given the general intractability of teacher behavior, not a very likely source of change at that. The more basic lesson it underscores should by now be a measurement commonplace--that the stability of any context, just like the reliability of its assessment, is an open empirical question, that the generalizability of a measure from one point in time to another requires recurrent response consistencies.

II. The context of salient traits and characteristics comprising the child's effective personality and intellectual makeup is most directly assayed through a strategy of multivariant measurement and analysis. That is, rather than measuring a single characteristic in isolation or even a collection of separate characteristics, one should assess and interpret multiple characteristics in relation to each other, using score or factor profiles or other forms of comparative and moderator analysis. Score interpretations should take into account evidence of interactive or moderator effects--that is, a high score for a particular characteristic may have a different meaning or different implications for individuals scoring high as opposed to low on a second characteristic or for individuals displaying a particular pattern of scores over a set of characteristics. Thus, the educational implications of a low score on a general information test may be quite different for a child who achieved moderately well on a variety of measures of problem solving and cognitive functioning as opposed to a child who performed poorly on those tasks. Or a consistent pattern of moderate to low performances on cognitive tasks might be interpreted somewhat differently

if accompanied by an extremely low score for memory or recall as opposed to a moderate or average score.

In the construction of comprehensive assessment batteries for children, emphasis is understandably given to dimensions of intellectual attainment, cognitive functioning, and sometimes even creative process, for these are closely attuned to major educational and social objectives. Less time is typically alloted to the assessment of affective dimensions, not because they lack educational or social relevance, but primarily because of difficulty in developing valid and efficient measures in the affective domain. Yet it is just such affective variables of motivation and interest and coping that provide the critical personal context necessary for drawing valid inferences about process or competency from cognitive test performance.

Given the interpretative importance of these affective variables, a provisional attempt has been made to assess them in the CIRCUS battery by turning once again to teachers' judgments. However, rather than asking teachers to make the kind of high-level inferences required to rate such chacteristics as aggressiveness or achievement motivation, with all the inherent biases entailed by such value-laden content, they are instead asked to rate each child in connection with a variety of activities. These activities, which include physical, motor, academic, language, role playing, fantasy, and artistic behaviors, are rated with respect to frequency of occurrence, degree of complexity, the creativity and imagination displayed, the amount of help or direction typically sought from adults, and the degree to which the child usually engages in the activity alone. If these ratings are sufficiently discriminating across children and display individual variability

across activities, then this activities inventory approach may provide serviceable measures of interests and of preferred or habitual coping styles in young children.

III. The context of the measurement process itself is most usefully assessed not so much by documenting objective characteristics of the tasks, the tester, and the situation as by recording the child's stylistic reactions to them. This is usually accomplished, following the lead of Hertzig et al. (1968), by means of direct tester or teacher observations of the child's stylistic responses to the cognitive demands or adaptive requirements of the measurement tasks. These ratings, which may be made separately for each task or a representative selection of tasks or globally for the battery as a whole, typically include judgments of such aspects of the child's responsiveness as the degree to which he asked for help, refused or indicated reluctance to work on tasks, expressed enjoyment or amusement over particular content, indicated he didn't know answers, indicated a desire to stop, appeared to respond "at random," appeared to weigh alternatives carefully, and spoke about or attended to unrelated objects or events. By relating stylistic consistencies in test responsiveness to patterns of test performance, the validity of test interpretation is likely to be improved, regardless of whether these response styles are transient and specific to particular tasks or situations or are more generally characteristic of the test taking behavior of the subject.

From this discussion it would appear that the major approach to the assessment of context is observational, that it is difficult to avoid the intrusion of human judgment in the measurement process. Although at this stage of the art, this may be true, it is not a critical issue to be

emphasized here. The important point is not that the assessment of context is inherently observational, but that it is inherently analytical. Dimensions of context are important because their potential interactive and moderator effects may differentially influence individual behavior. Hence, the descriptive measurement of a variety of dimensions, however salient or pervasive, is not enough for a true assessment of context--in addition the interactions and moderated relationships must be assessed or revealed analytically. For a comprehensive program of measurement to deal meaningfully with the assessment of context, then, it must include provision for multivariate analysis and for the display, reporting, and interpretation of interactive and moderated relationships.